
IBM PROJECT

RESEARCH AGENT

Presented By:
Saranga Pradeep
Ace Engineering College
Cse (Data Science)

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Researchers often spend a lot of time searching for papers, summarizing content, managing citations, and drafting reports. These tasks are repetitive and slow down the research process. There is a need for an AI system that can help with these activities. A Research Agent can understand questions, find relevant papers, create summaries, organize references, and even help write research sections. This makes research faster, more accurate, and more efficient for both academic and industrial use.

PROPOSED SOLUTION

- The proposed system aims to solve the challenges researchers face in managing academic tasks like literature search, paper summarization, citation organization, and research drafting. To address this, we have developed an AI-based Research Agent using IBM Cloud Lite services and IBM Granite models.
- **Research Query Understanding::**
 - The agent accepts natural language input from users (e.g., research questions or keywords). Using IBM Granite LLMs, it interprets the query context and prepares it for search and analysis.
- **Document Search & Knowledge Retrieval::**
 - It connects to open-access research repositories such as Semantic Scholar, arXiv, and CrossRef APIs to retrieve relevant academic papers or metadata. Retrieved results are filtered and ranked for relevance.
- **Summarization & Content Analysis::**
 - The Research Agent uses IBM Watsonx and Granite models to summarize uploaded or fetched research papers. It can generate abstract-style summaries, key points, or section-wise breakdowns (e.g., Introduction, Methodology, Conclusion).
- **Deployment:**
 - The solution is hosted using IBM Cloud services. Backend is developed in Python and deployed via IBM Code Engine or IBM Cloud Functions for serverless scalability. Files and documents are stored using IBM Cloud Object Storage.
- **Evaluation:**
 - The system's responses are evaluated based on relevance, summarization quality, and citation accuracy. User feedback is used to improve prompt engineering and fine-tune the response logic.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology used to design, develop, and deploy the Research Agent system on IBM Cloud. This includes the infrastructure, tools, frameworks, and AI models integrated to achieve the desired functionality.

■ System requirements

- IBM Cloud Lite Account
- IBM Watsonx.ai and IBM Granite Model Access
- IBM Cloud Object Storage (for file and document storage)

■ Library required to build the model

- IBM Watsonx & Granite – for NLP tasks like summarization, question answering, and hypothesis generation
- Markdown/HTML parsers – to render formatted summaries or citation outputs on frontend

ALGORITHM & DEPLOYMENT

■ Algorithm Selection:

- The system uses IBM's **Granite large language models** (LLMs) through Watsonx.ai for natural language processing tasks such as summarization, question answering, citation generation, and hypothesis suggestion. These models were selected for their accuracy in understanding academic language and generating coherent, structured outputs. Unlike traditional ML models, Granite LLMs are pre-trained on vast text corpora and perform well with zero- or few-shot prompting.

■ Data Input:

- The inputs to the system include:
- User queries (natural language research questions)
- Uploaded research papers (PDF documents converted to text)

■ Training Process:

- Extracted text from papers is chunked and embedded using Granite or other embedding models.
- Vector similarity search is performed to retrieve relevant content based on user queries.
- Retrieved content is passed as context to prompt templates in Granite for summarization, Q&A, or citation generation.

■ Prediction Process:

- Deployed via **IBM Code Engine** or **IBM Cloud Functions** for scalability. File storage is handled using **IBM Cloud Object Storage**
- Watsonx.ai is used to call Granite LLMs via prompt chaining or REST APIs. The system can be accessed through a simple web interface or chatbot UI.

RESULT

IBM watsonx

Upgrade ? 🔔

Pradeep Saranga's Account ▾ Dallas ▾ PS

Deployment spaces / Research Agent / Research Agent /


Research Agent ✔ Deployed Online

API reference Test **Preview**

New chat +

P You 10:10 PM

Find recent papers on AI in healthcare.

 Research Agent 10:10 PM

Here are some recent papers on AI in healthcare:

1. "Artificial intelligence in healthcare: transforming the practice of medicine" (2023) - This paper summarizes recent breakthroughs in the application of AI in healthcare, including the use of automated AI algorithms for diabetic retinopathy.


2. "Revolutionizing healthcare: the role of artificial intelligence in clinical practice" (2023) - This paper discusses the potential of AI to transform healthcare, including its use in developing diagnostic tools and personalized treatment plans.

3. "A Review of the Role of Artificial Intelligence in Healthcare" (2023) - This paper reviews the current state of AI in healthcare, including its applications in medical research, diagnosis, and treatment.


4. "Journal of Medical Artificial Intelligence" (2023) - This journal publishes articles on the latest research and innovations in medical AI, including its applications in healthcare.

Type something

About this deployment ×

Name 


Research Agent



Description 


The Research Agent is an AI tool that finds, summarizes, and organizes academic content using NLP.


Deployment Details


Deployment ID: `c677940b-712f-45...`

Serving name:  No serving name.


Software specification: 
`runtime-24.1-py3.11` 


Hardware specification: 
Extra extra small: 1 CPU and 2 GB RAM

Copies: 
1

Tags 

wx-agent

Associated asset 

 [Research Agent](#)

edunet
foundation

RESULT

The screenshot displays the IBM watsonx Agent Lab interface. The top navigation bar includes the IBM watsonx logo, an 'Upgrade' button, a user profile for 'Pradeep Saranga's Account', and location settings for 'Dallas'. The main header shows the project path 'Projects / ResearchAgent / Agent Lab' and buttons for 'Share feedback', 'Unsaved', 'New agent +', and 'Deploy'.

The interface is divided into two main panels:

- Build Panel:** This panel is used to configure the agent. It includes a 'Setup' section with fields for 'Name' (set to 'Research Agent') and 'Description' (set to 'The Research Agent is an AI tool that finds, summarizes, and organizes academic content using NLP.'). There is also an 'Icon' section with a placeholder image and a 'Quick start questions' section with a text input field.
- Agent preview Panel:** This panel shows a simulated conversation with the agent. The conversation starts with a user message 'hi' at 08:43 PM, followed by an agent response: 'Hello! How can I assist you with your research today? Do you have a specific question, topic, or task you'd like help with?'. The user then asks 'i want to help in the research paper of the Phishing url ml project' at 08:45 PM. The agent responds: 'That sounds like an interesting project! A Phishing URL ML project aims to develop a machine learning model that can detect and classify phishing URLs. Here are some potential ways I can assist you with your research paper: 1. Literature Review: I can help you find and summarize relevant academic papers related to'. The interface also shows a 'How did I get this answer?' section with a dropdown menu set to '1: Generating final answer'.

The bottom of the screenshot shows a Windows taskbar with the date '03-08-2025' and time '20:47'.

RESULT

The screenshot displays the IBM watsonx Agent Lab interface. The top navigation bar includes the IBM watsonx logo, an 'Upgrade' button, a user profile for 'Pradeep Saranga's Account' in 'Dallas', and a 'PS' status indicator. The main header shows the project path 'Projects / ResearchAgent / Agent Lab' and buttons for 'Share feedback', 'Unsaved', 'New agent +', and 'Deploy'.

The interface is divided into two main sections: 'Build' and 'Agent preview'.

Build Section:

- Name:** Research Agent
- Description:** The Research Agent is an AI tool that finds, summarizes, and organizes academic content using NLP. (98/100 characters)
- Icon:** A placeholder image showing a robot head and a magnifying glass over a document.
- Quick start questions:** A list of questions for users to select, including 'What are the key findings of this article?'.

Agent preview Section:

Research Agent 08:45 PM

That sounds like an interesting project! A Phishing URL ML project aims to develop a machine learning model that can detect and classify phishing URLs. Here are some potential ways I can assist you with your research paper:

1. Literature Review: I can help you find and summarize relevant academic papers related to phishing URL detection, machine learning, and cybersecurity.
2. Data Collection: If you need help collecting a dataset of phishing and legitimate URLs, I can assist you in finding publicly available datasets or provide guidance on how to collect your own data.
3. Feature Engineering: I can help you identify and extract relevant features from the URLs that can be used to train a machine learning model.
4. Model Development: I can provide guidance on selecting a suitable machine learning algorithm and hyperparameters for your model.
5. Results Analysis: Once you have trained and tested your model, I can help you analyze the results and provide insights on how to improve the model's performance.
6. Writing and Editing: I can assist you in writing and editing your research paper, ensuring that it is well-structured, clear, and free of errors.

To get started, could you please provide more details about your project, such as:

- What is the specific goal of your project (e.g., detecting phishing URLs in real-time, classifying URLs as phishing or legitimate)?

Type something...

RESULT

The screenshot displays the IBM Watsonx Agent Lab interface. The top navigation bar includes the IBM Watsonx logo, a user profile for Pradeep Saranga, and account details for Dallas. The main content area is divided into two panels: 'Build' and 'Agent preview'.

Build Panel:

- Name:** Research Agent
- Description:** The Research Agent is an AI tool that finds, summarizes, and organizes academic content using NLP.
- Icon:** A placeholder image showing a robot head and a magnifying glass over a document.
- Quick start questions:** A section for adding questions, with one example: "What are the key findings of this article?"

Agent preview Panel:

You: 08:46 PM
What features are you extracting from the URLs?

Research Agent 08:46 PM
For a Phishing URL ML project, some common features that can be extracted from URLs include:

- 1. URL Length:** Phishing URLs are often longer than legitimate URLs to accommodate the phishing payload.
- 2. URL Depth:** Phishing URLs may have a deeper directory structure to make them appear more legitimate.
- 3. Number of Dots:** Phishing URLs may contain more dots (.) to create subdomains that appear legitimate.
- 4. Number of Hyphens:** Phishing URLs may contain more hyphens (-) to create subdomains that appear legitimate.
- 5. Number of Slashes:** Phishing URLs may contain more slashes (/) to create a deeper directory structure.
- 6. Presence of Suspicious Words:** Phishing URLs may contain words like "login", "password", "update", etc. that are commonly used in phishing attacks.
- 7. Presence of Special Characters:** Phishing URLs may contain special characters like @, _, -, etc. that are not typically found in legitimate URLs.
- 8. Domain Age:** Phishing URLs may have a younger domain age than legitimate URLs.
- 9. Domain Reputation:** Phishing URLs may have a poor domain reputation based on blacklists, whitelists, or other reputation systems.

Type something...

RESULT

Deployment spaces /

Research Agent

Overview Assets **Deployments** Jobs Manage

| Name | Type | Status | Asset | Asset type | Tags | Last modified |
|----------------|--------|----------|----------------|------------|----------|---|
| Research Agent | Online | Deployed | Research Agent | Ai service | wx-agent | 42 seconds ago Pradeep Saranga (You) |

Items per page: 20 1-1 of 1 items 1 of 1 pages

29°C Mostly cloudy Search ENG IN 21:11 03-08-2025

CONCLUSION

- The Research Agent helps make academic work easier by finding papers, summarizing them, managing citations, and drafting research content. It uses IBM Cloud and Granite models to provide fast and accurate support. The system saves time and improves the quality of research. Some challenges faced were in getting accurate results and handling PDF files, but they were solved using proper tools and prompts. Overall, the Research Agent is a useful tool for students and researchers, and it can be improved further by adding more features in the future.

FUTURE SCOPE

- Enable **plagiarism detection** and originality checking for research content.
- Recommend suitable **journals or conferences** for paper submission.
- Introduce **voice input** and **voice response** for hands-free interaction.
- Add **real-time chat interface** for conversational research assistance.
- Allow **collaborative project spaces** for teams working on the same research.
- Include **research deadline reminders** and progress tracking features.
- Expand citation tools to support **LaTeX, BibTeX, and Zotero integration**.
- Build a **mobile-friendly version** for access on the go.

REFERENCES

- IBM Watsonx.ai Documentation – <https://www.ibm.com/cloud/watsonx>
- IBM Granite Foundation Models – <https://www.ibm.com/blogs/research/2023/12/ibm-granite-models/>
- PyMuPDF (fitz) Library – <https://pymupdf.readthedocs.io/>

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



PRADEEP SARANGA

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 22, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/e7e2721d-5af3-4e30-97f4-0ecf21feb827>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



PRADEEP SARANGA

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution




Issued on: Jul 22, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/05bc0fdf-6ce2-41dd-846a-d9cf1e755e64>



IBM CERTIFICATIONS

IBM SkillsBuild Completion Certificate



This certificate is presented to
PRADEEP SARANGA
for the completion of
**Lab: Retrieval Augmented Generation with
LangChain**
(ALM-COURSE_3824998)
According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins



THANK YOU